

RESEARCH ARTICLE

The enigma of the near-symmetry of proteins: Domain swapping

Maayan Bonjack-Shterengartz, David Avnir*

Institute of Chemistry and the Lise Meitner Minerva Center for Computational Quantum Chemistry, The Hebrew University of Jerusalem, Jerusalem, Israel

* david.avnir@mail.huji.ac.il



Abstract

The majority of proteins form oligomers which have rotational symmetry. Literature has suggested many functional advantages that the symmetric packing offers. Yet, despite these advantages, the vast majority of protein oligomers are only *nearly* symmetric. A key question in the field of proteins structure is therefore, if symmetry is so advantageous, why do oligomers settle for aggregates that do not maximize that structural property? The answer to that question is apparently multi-parametric, and involves distortions at the interaction zones of the monomer units of the oligomer in order to minimize the free energy, the dynamics of the protein, the effects of surroundings parameters, and the mechanism of oligomerization. The study of this problem is in its infancy: Only the first parameter has been explored so far. Here we focus on the last parameter—the mechanism of formation. To test this effect we have selected to focus on the domain swapping mechanism of oligomerization, by which oligomers form in a mechanism that swaps identical portions of monomeric units, resulting in an interwoven oligomer. We are using continuous symmetry measures to analyze in detail the oligomer formed by this mechanism, and found, that without exception, in all analyzed cases, perfect symmetry is given away, and we are able to identify that the main burden of distortion lies in the hinge regions that connect the swapped portions. We show that the continuous symmetry analysis method clearly identifies the hinge region of swapped domain proteins—considered to be a non-trivial task. We corroborate our conclusion about the central role of the hinge region in affecting the symmetry of the oligomers, by a special probability analysis developed particularly for that purpose.

OPEN ACCESS

Citation: Bonjack-Shterengartz M, Avnir D (2017) The enigma of the near-symmetry of proteins: Domain swapping. PLoS ONE 12(7): e0180030. <https://doi.org/10.1371/journal.pone.0180030>

Editor: Dinesh Gupta, International Centre for Genetic Engineering and Biotechnology, INDIA

Received: October 10, 2016

Accepted: June 8, 2017

Published: July 14, 2017

Copyright: © 2017 Bonjack-Shterengartz, Avnir. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The study was supported by the Israel Science Foundation Grant (grant #411/15). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The abundance of chiral rotational symmetry in protein oligomers[1–9] raises an interesting question: On one hand the list of advantages of this symmetrization is comprehensive and includes increasing the protein stability, avoiding excessive aggregation, enhancing of coding efficiency, reducing of synthetic errors, and inducing efficient cooperative regulation[1–5]. On the other hand, despite these advantages, we have shown recently[10] that perfect symmetry in proteins is rare: many oligomers which are built not only from similar (hetero-oligomers) building units but even from identical (homo-oligomers) deviate from ideal, perfect symmetry

to some degree. This deviation is always detectable and measurable, and is beyond experimental uncertainty. What then is the origin of symmetry deviation that does not allow oligomers to maximize the symmetrization advantages? Recently we have proposed [10] that parameters which may be relevant for this question mark are: the minimization of the enthalpy of the interactions of the amino-acid units at the contact zones of the oligomeric subunits, which require giving away symmetry in order to attain that optimization (dealt with and proven in ref. [10]); relaxing the high entropic cost of maintaining perfect symmetry by increasing the number of possible microscopic conformations states of the protein; the operation of the property of any dynamic process that shifts objects away from symmetry; and the effects of the surrounding environment of the oligomer (solvent, crystal neighbors, the hydration shell), which may stabilize a distorted structure.

Here we explore the mechanism of the oligomerization as a potential source for symmetry deviation in protein oligomers. The rationale behind assuming that the formation of an oligomer may affect its symmetry is that the protein structure may reflect steps it underwent during its formation. For example, when the oligomer consists of at least three monomers, the mechanism of oligomerization is prone to be a sequential [11,12] (and not, at least in part, concerted), a route which may lead to de-symmetrization, because the first step is dimerization, and the next one is an interaction of a monomer with a dimer. In dimeric proteins—which are the main focus of this report—as well as in higher oligomers, the symmetry may be affected by the specific nascent stage after translation of all or part of the monomeric unit chains, at which association to form the dimer commences—it may take place either only after full completion of the monomer synthesis, or at an earlier stage [1,11–13].

A particularly interesting mechanism of oligomerization which belongs to the latter option is domain swapping. The general idea of that proposed mechanism is that when two (or more) monomeric units assemble, they do so not by a simple aggregation process, but by aggregation that is accompanied by mixing or exchange of identical structural elements of the subunits [14–17]. In the swapping mechanism that mixing is carried out by exchanging (swapping) identical structural domains, so that two or more identical protein molecules form an intertwined oligomer, as shown in Fig 1. The resulting oligomer formed by this mechanism consists

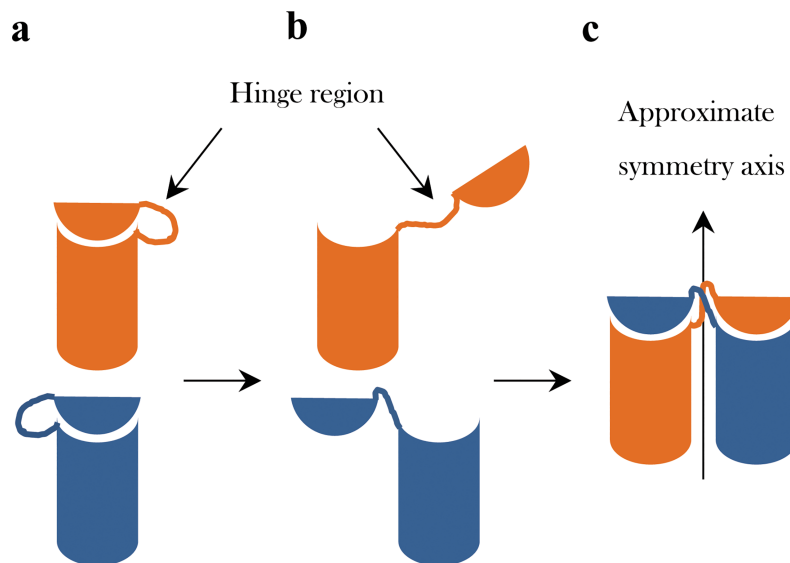


Fig 1. The domain swapping mechanism, demonstrated on the formation of a dimeric oligomer. (a) Two monomers with their folded potential hinge regions. (b) The monomers with their open hinge regions. (c) The dimerization, leading to the domain-swapped oligomer.

<https://doi.org/10.1371/journal.pone.0180030.g001>

of subunits with the same structure as of the original monomer, except for the linking segments known as the *hinge regions* which connect the swapped domains (the secondary minor region) with the rest of the structure (the secondary major region). This oligomerization mechanism has been proposed for a wide range of proteins[15,18–24] where the size and nature of the swapped domains vary and may be as small as one secondary structural element or as large as a significant portion of the whole protein molecule. Likewise, the hinge region may be as small as consisting of three amino acids, but is it rarely larger than 15 amino-acids in length[21]. The majority of the oligomers formed by the swapping process display C_n symmetry. This cyclic symmetry group contains a single axis of rotational symmetry, characterizing a protein with a quaternary structure of n subunits arranged in a ring, and which are related by an n -fold axis. The most prevalent ones are of C_2 -symmetry[1] (which describes a half-turn symmetry), that is, dimers, which are therefore the focus of this report.

We report here our finding that, in agreement with our general observation cited above [10], that many dimers which are categorized as swapped-domain oligomers deviate from perfect symmetry. This observation has led us to investigate the hypothesis that the cause of this general symmetry deviation is related to the swapping mechanism, and particularly to the resulting linking hinges regions of the sub-units. This is so because the hinge region in each of the monomeric units is the only region that changes its secondary structure drastically when this mechanism operates: Often the change is from a folded minor-major region link within the monomeric state to an extended conformation link of these regions (Fig 1). If this is indeed the case then symmetry analysis which focuses on the symmetry relation of the two hinge regions (one in each subunit) may highlight them as carrying most or at least some of the distortive burden of these oligomers. In this report we show that, indeed, symmetry analysis identifies faithfully the hinge regions as significant symmetry distorted portions of the oligomers. It is also interesting to note in this context that in most cases of domain swapped proteins, the hinge region is located at or very close to the near- C_2 axis (Fig 1).

We recall that supporting evidence for the swapping mechanism is not trivial, and that the full and detailed molecular swapping mechanism and its exact energetic aspects are still under development. From that point of view, the symmetry analysis presented below may also serve as supporting evidence for a swapping mechanism, when such is proposed. Propositions of domain swapping have been categorized as follows[15]: ‘*Bona fide* domain swapping’ proteins are such that their monomeric form is known; ‘Quasi-domain swapping’ proteins are such that a monomeric homologue is known; and ‘candidates for domain swapping’, which are proteins for which structural information of their monomer or monomeric homologue form is not available. In the last two decades several methods were developed[14,19–21,25–27] in order to address the question of whether a protein was formed by domain swapping mechanism and in order to identify the exact location and size of the hinge region in a protein oligomer suspected to be formed by that mechanism. The main method in this field was developed by Eisenberg and his co-workers[14] and is suitable for *bona-fide* domain swapping and quasi-domain swapping proteins, and utilizes a superimposability test between the hinge regions in the monomer and the dimer. See also instance 20 and 21 for more improved versions of Eisenberg’s method. In cases of the third category—candidates of domain swapping proteins—the hinge loop region has been looked-after by several methods such as direct inspection of the protein’s crystallographic structure[19], or by the determination of the global minimum of the compactness profile of the oligomer[25]; of course, these methods are also suitable for the first two categories.

As was described above, tools for screening of domain swapped proteins already exist and the main contribution of the CSM analysis are for cases of uncertainty about the relevance of the domain swapping mechanism, for strengthening (or excluding) this proposed mechanism,

and for accurately determining the protein hinge region. In the following sections, we first present the symmetry analysis that we developed in order to address proteins with proposed domain swapping mechanism; this method identifies the hinge region of swapped domain proteins with no need of structural information on the monomeric form of the non-swapped protein. We then provide an overall picture of the symmetry analysis results and their generality, include detailed investigation of several cases, and discuss the influence of the domain swapping mechanism on symmetry distortions of the whole oligomer, proving, we believe, that the formation of an oligomer may have profound effect on the resulting degree of symmetry.

Methods

The computational tools

The main focus of this study is the symmetry of proteins. The voluminous literature on this structural property of proteins has been limited by a qualitative descriptive language (“near-symmetry”, “approximate symmetry”, etc.) [1–4,7,28]. A quantitative approach which answers questions such as, ‘what is the degree of symmetry of an approximate-symmetry protein’, and, ‘by how much is one pair of hinges more or less C_2 -distorted than another pair’ would allow to transfer the whole analysis and discussion to measurable facts. Thus, all of the symmetry analyses in this report are based on the Continuous Symmetry Measure (CSM) [29,30], a method for quantifying the degree of symmetry of a given object. According to the CSM approach, the G -symmetry point group content of an object is the minimal distance between two objects: an original structure and a G -symmetric structure, \bar{Q}^{sym} , which consists of the same atoms and connectivity and is the closest to the original distorted structure. This minimal distance of the object’s vertices from the desired G -symmetry defined the measure $S(G)$:

$$S(G) = \frac{100}{d^2} \sum_{i=1}^N |\bar{Q}_i - \bar{Q}_i^{sym}|^2, \quad (1)$$

where \bar{Q}_i are the coordinates of the i^{th} atom of the original studied molecule, \bar{Q}_i^{sym} are the coordinates of the i^{th} atom of the nearest structure which has the desired symmetry, the denominator is the root mean square size normalization factor of the original centered structure ($d = \sqrt{\sum_{i=1}^N |\bar{Q}_i|^2}$), and N is the number of analyzed atoms in the structure (see full details in [10,31]). It should be emphasized that this measure is inherently different than the rmsd analysis of the degree of similarity—the rmsd analysis does not evaluate the symmetry itself as a structural parameter, which is the key issue of this report. The range of the symmetry measure is $0 \leq S(G) \leq 1$ and it is expanded by a factor of 100 for convenience ($0 \leq S(G) \leq 100$). If a structure is of perfect G -symmetry, then $S(G) = 0$ and as the structure distorts from the perfect symmetry, $S(G)$ increases. $S(G)$ is a special distance function in that the nearest \bar{Q}_i^{sym} is usually not known a-priori, but is determined by a minimization protocol described in detail in previous publications [29,32,33]. The measure is a global parameter, and therefore allows the comparison of various structures and various symmetries on the same scale. For alternative symmetry and chirality measures see, e.g., ref.’s [34] and [35].

In a previous study [10] we have introduced specific CSM computational tools for the evaluation of the symmetry content, $S(G)$, of proteins, two of which are relevant for this report: The “symmetry analysis of fragments” and the “local symmetry analysis”. The fragments analysis, as the name implies, focuses on symmetry relations of specific portions of the protein structure. This analysis might reveal, on one hand, which regions in the protein carry the burden of the deviation, and on the other hand, which are barely deviating from perfect symmetry. The analyzed fragments can be as small as symmetry related atoms, but we found that the relevant

minimal, useful unit in the context of proteins is the individual amino-acid; when this is used we refer to the analysis as a *local symmetry analysis*, which is a high-resolution tool: A CSM calculation is carried out on each pair of symmetry-matched amino acids within an oligomer, one amino-acid from each monomer. Each such calculation provides a local CSM values. This local symmetry analysis gives at a glance the relative deviations from symmetry within the oligomer structure, and specifically reveals which pairs of amino-acids are the most distorted ones in the structure. Detailed examples below, clarify it further.

The analyzed proteins data

The selection of domain swapping protein structures for analysis was based on the datasets of Eisenberg[15] and of Huang[19] and on "3DSwap Knowledgebase of 3D domain swapping in proteins" database[36]. The coordinates of the analyzed proteins (\bar{Q}_i in Eq (1)) were taken from the crystallographic Protein Data Bank (PDB)[37]. All PDB entries in which the subunits are related by crystallographic symmetry are excluded from our data set. Therefore, we did not use any data in the database or in the literature mentioned above which was derived by placing only one sub-unit in the asymmetric unit and then assuming complete symmetry (these structures are by definition of $S(G) = 0$ value); the only crystallographic asymmetric units taken contain the full oligomer in the asymmetric unit.

Results and discussion

The CSM spectrum analysis

For the analysis of the rotational symmetry of the hinge regions—a pair of hinge regions in the case of C_2 -symmetry—we developed the following extension of the symmetry analysis of fragments described above: A segment of h amino-acids is selected; h is defined as the *size of analysis ruler*. Then, (see Fig 2), starting with the 1st amino-acid in the polypeptide chain of the monomer, the $S(C_2)$ value of the first C_2 -symmetry-related segment - 1st- h^{th} amino-acids segments-pair—is calculated (without H atoms), and a first CSM value is obtained. The ruler is moved then by a one amino-acid step to the second segment—2nd-($h+1$)th amino-acids—and a second CSM value is calculated. The procedure is repeated one amino-acid after the other with the “running ruler” until (and including) the final segment of length h is reached. A total of $N = n - h + 1$ (where n is the number of amino-acids in the subunit) segments and their associated CSM values are obtained. A *CSM spectrum* is then plotted (Fig 3) in which the CSM value ($S(C_2)$) of the i -th segment (y-axis) is presented as a function of the position, n_i , of the first amino acid in that segment (x-axis). The main idea is that zones in the protein which deviate more than their neighboring zones, should appear as peaks of high $S(C_2)$ values. The running ruler can be of any size: As short as one amino-acid (“*local symmetry analysis*”), or as long as and the whole size of the polypeptide chain (“*all-atoms symmetry analysis of whole protein oligomer*” (see Ref. [10])). We have sampled different sizes of the ruler, and found that if nothing is known about the hinge in a suspected oligomer, one should use a ruler of size 10, and if a proposition exists about the size of a suspected hinge, one should test first that size as a running ruler (a case where we start with that proposed size, but then find a different size which is better will later be shown).

The generality of the symmetry distortion of the hinge range pairs. We have carried out this CSM spectrum analysis on various protein structures suggested to be formed by a domain swapping mechanism. All-and-all, we have used 40 arbitrarily selected protein structures. For all proteins, the CSM spectra were obtained by the running-ruler method, and the spectra analyzed. A typical CSM spectrum is displayed in Fig 3B for RNase A N-terminal

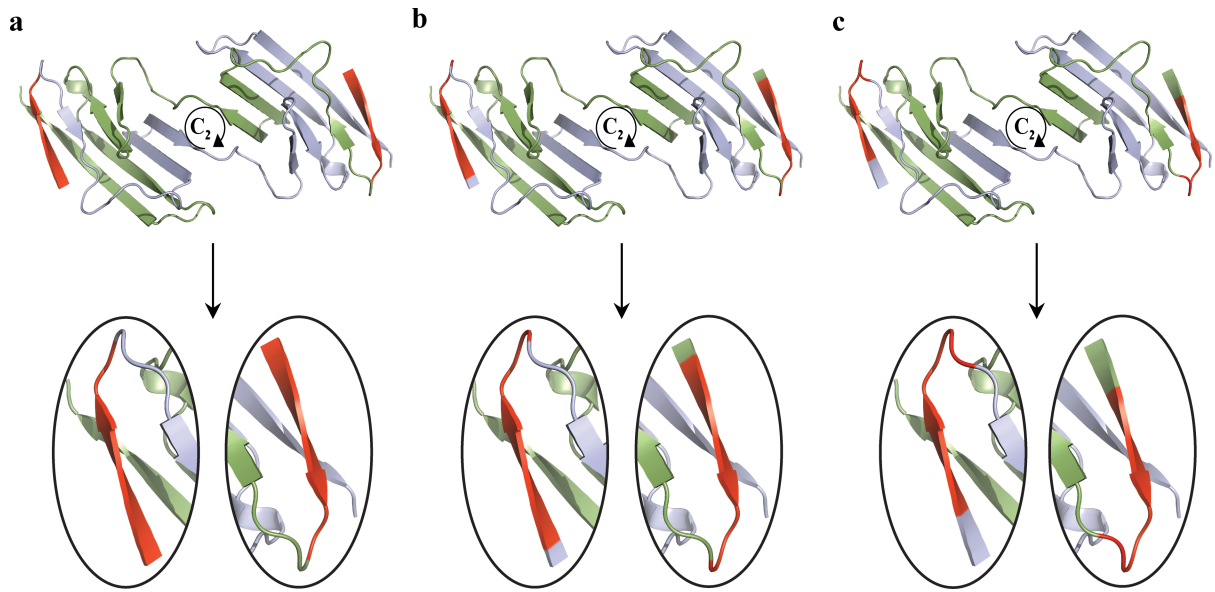


Fig 2. The running ruler method demonstrated on the engineered N-terminal domain of CD2 protein (PDB code: 1A64), starting from the N-terminal; running ruler size (red): seven amino-acids ($h = 7$). (a) The first segment, 1st-7th amino acids segments-pairs. (b) The second segment, 2nd-8th amino acids segments-pairs. (c) The third segment, 3rd-9th amino acids segments-pairs.

<https://doi.org/10.1371/journal.pone.0180030.g002>

swapped dimer (PDB code: 1A2W), the structure of which is shown in Fig 3A. The general feature seen in Fig 3B is a sharp peak at the amino-acids positions range of 16–23, which very closely coincides with the amino-acids range originally suggested, namely, 15–22 –indicated in Fig 3A. This region is significantly more symmetry-distorted compared to any other segment in the protein, that is, it carries most of the burden of the symmetry. Let us assume that nothing is known about the hinge of this oligomer; we then have to use a ruler size of 10, which is displayed in Fig 3C. The CSM spectrum still identifies this region as the hinge region, but with less accuracy (the range now is 16–25). Table 1 summarizes similar observations made for proteins which belong to the *bona-fide* domain swapping and quasi-domain swapping categories, and that their hinge region locations were determined by Eisenberg et al.[15]; the related CSM

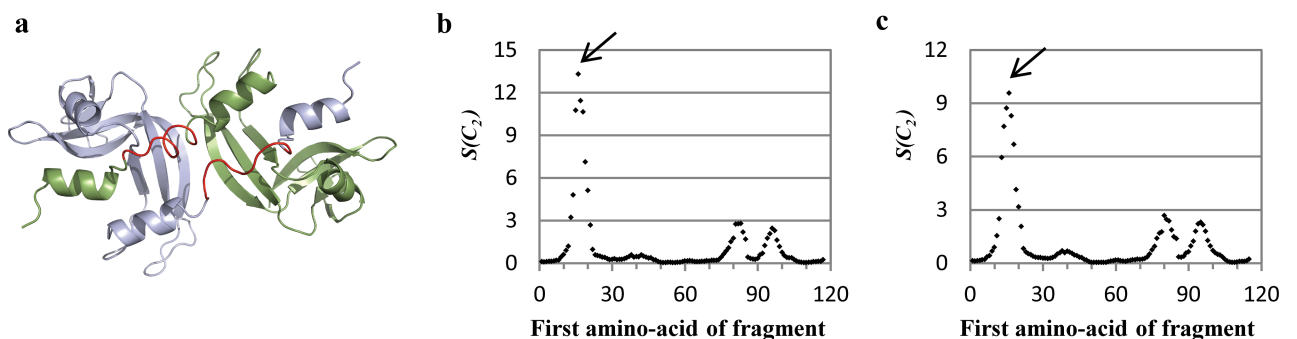


Fig 3. Running ruler symmetry analysis applied on RNase A N-terminal swapped dimer (PDB code: 1A2W). (a) Cartoon representation of the protein. Each subunit is indicated by a different color (blue and green), and the originally proposed hinge region is colored red. (b) CSM spectrum of the protein, the ruler size is as the length of the hinge region (8 amino acids). The black arrow indicates the hinge region. (c) CSM spectrum of the protein with a ruler size of 10 amino acids. The black arrow indicates the hinge region. For data source see ref. [38].

<https://doi.org/10.1371/journal.pone.0180030.g003>

Table 1. Hinge region location.

PDB ID	Hinge region from the symmetry analysis	Literature hinge region[15]	Shift between the methods
1WWA	295–297	297–299	2
2SPC	70–73	72–75	2
1LMK	122–126	123–127	1
1JS0	111–114	112–115	1
1HUL	81–88	82–89	1
1CDC	43–49	44–50	1
1OBP	121–124	121–124	0
1H8X	17–24	16–23	-1
1FYR	122–124	121–123	-1
1A2W	16–23	15–22	-1
1AOJ	36–41	34–39	-2
1BSR	17–24	15–22	-2
1BLB	81–89	79–87	-2
1HT9	41–50	38–47	-3
1BH5	24–36	20–32	-4

<https://doi.org/10.1371/journal.pone.0180030.t001>

spectra are collected in Figs 4–6. (in these spectra, one should consider the relative values of S (C₂) in each spectra rather than its absolute values. It is seen from the Table that our method identifies hinge regions in all cases, and that in general they overlap well, with minor shifts of 1–2 residues, compared to the original propositions. Even the two last entries in Table 1 which display shifts of 3 and 4 amino acids, belong to large hinge regions, and represent overlaps of 7 and 9 amino-acids, respectively. *Without exception, in all proteins we analyzed, the hinge area appears as a peak, even in the third category of “candidates of domain swapping proteins”*—the generality is shown in Figs 5 and 6 and in S2 Fig. Thus, the formation route of the oligomer emerges as a key parameter in explaining its giving-up perfect symmetry. In the Probability analysis section we strengthen this conclusion with a statistical analysis, but some further comments on the data that can be elucidated from the CSM spectra is due first:

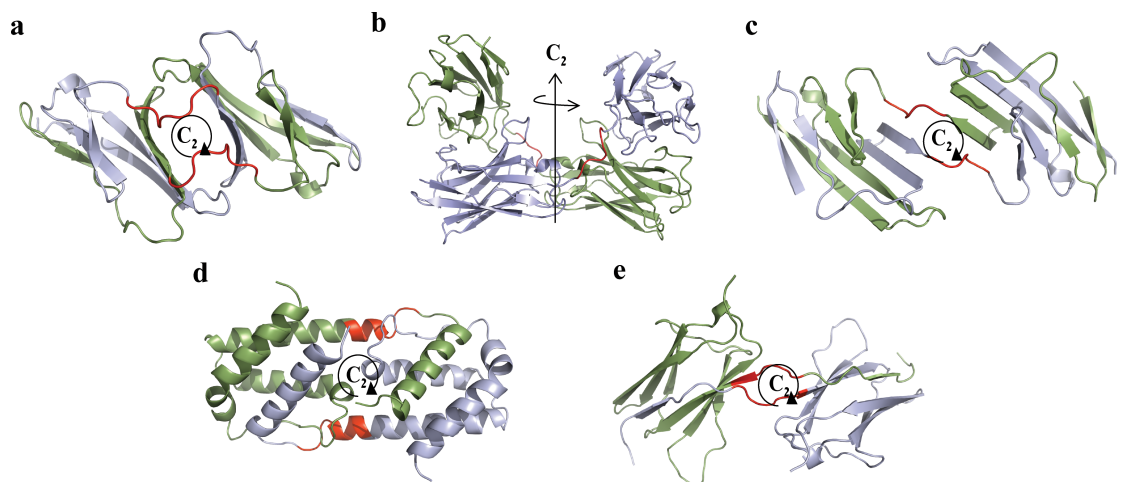


Fig 4. The proteins structures analyzed in Fig 5. Each subunit is indicated by a different color, and the originally proposed hinge region is colored red. (a) N-terminal domain of CD2 (PDB code: 1CDC), (b) Diabody (PDB code: 1LMK), (c) Engineered N-terminal domain of CD2 (PDB code: 1A64), (d) Interleukin-5 (IL-5, PDB code: 1HUL), (e) TrkA-d4 dimer (PDB code: 1WWA). For data sources see ref.’s [39–43].

<https://doi.org/10.1371/journal.pone.0180030.g004>

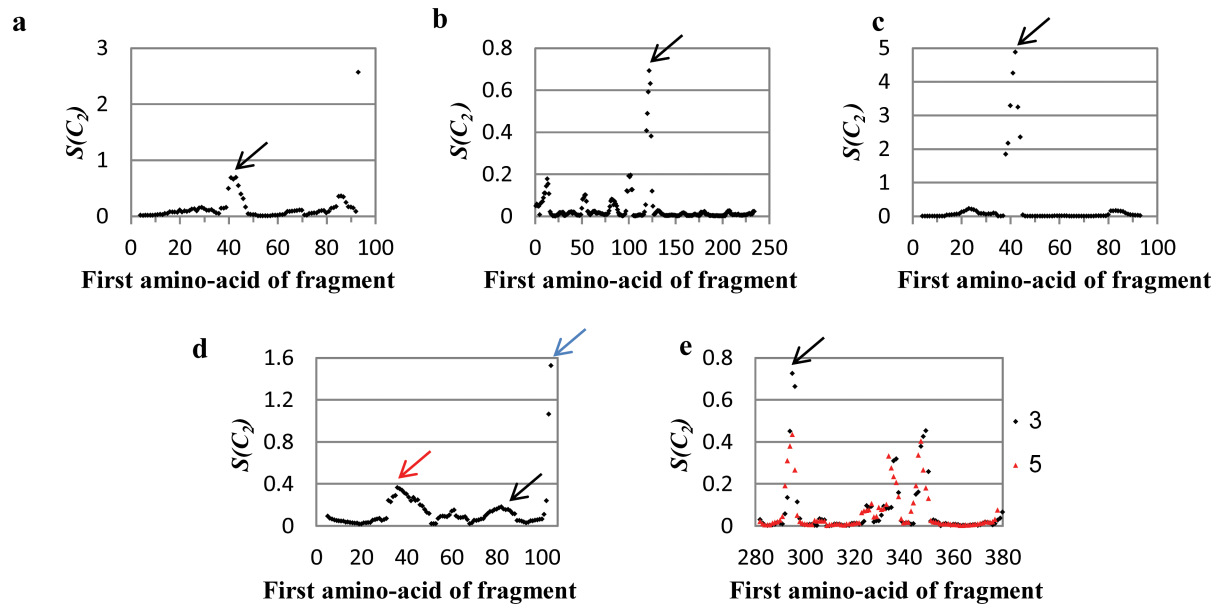


Fig 5. Running ruler symmetry analysis applied on proteins involved in 3D domain swapping. The black arrow indicates the hinge region; other colored arrows are explained in the text. (a) N-terminal domain of CD2, hinge region: 44–50, (b) Diabody, hinge region: 123–127, (c) Engineered N-terminal domain of CD2, hinge region: 44–50, (d) Interleukin-5 (IL-5), hinge region: 82–89, (e) TrkA-d4 dimer, hinge region: black– 297–299, red– 295–299. See Fig 4 for their PDB codes and cartoon representation and Table 1 for more information.

<https://doi.org/10.1371/journal.pone.0180030.g005>

Further comments on the CSM spectra. It is not necessary that the hinge region is the only portion of the protein which is C_2 -symmetry distorted, or that the hinge pair is the most symmetry-distorted region in the oligomer. For instance, let us look again at the CSM spectrum of the engineered N-terminal domain of CD2 (Fig 5C), which has seven amino-acids hinge loop, located at the amino-acids 44–50. The most distorted region in the structure, as indicated in the spectrum indeed points to the hinge segment at the amino-acids 42–48, with minor shift of two residues compared to the originally suggested hinge region (44–50, according to 3DSwap Knowledgebase of 3D domain swapping in proteins)[36]. This region is significantly different from the rest of any segment in the protein, from the symmetry point of view: it carries most of the burden of the symmetry deviation. It is also seen that the spectrum indicates additional distorted regions—two additional peaks at the 21–28 and 81–88 segments (and their counterparts in the second arm of the dimer). The origin of this distortion becomes clear upon careful examination of its 3D structure (Fig 7A): It is seen that the two segments are over-crowdedly very close to each other, and thus, to alleviate this disfavored situation, these segments give-up some of the mutual symmetry for better spatial alignment. It is thus evident that the CSM spectrum and the running ruler method can be used generally for analyzing structural features of proteins other than those originating from the swap mechanism.

Next, let us analyze a case where the hinge peak does exist, but is not the highest, specifically, interleukin-5, Fig 4D, which is a swapped domain protein with a hinge region of 8 amino-acids[15]. Applying the running ruler analysis on this protein creates a CSM spectrum with few peaks (Fig 5D), two of which are higher than the hinge region peak (81–88). The most distortive segment in the structure is at the C-terminal segment (indicated by a blue arrow). Such zones, of either N- or C- terminal segments, tend to distort from perfect symmetry. This is so because of the flexibility of the polypeptide chain termini. This observation is seen again in the C- terminal segment of the N-terminal domain of CD2 (Fig 5A). The second

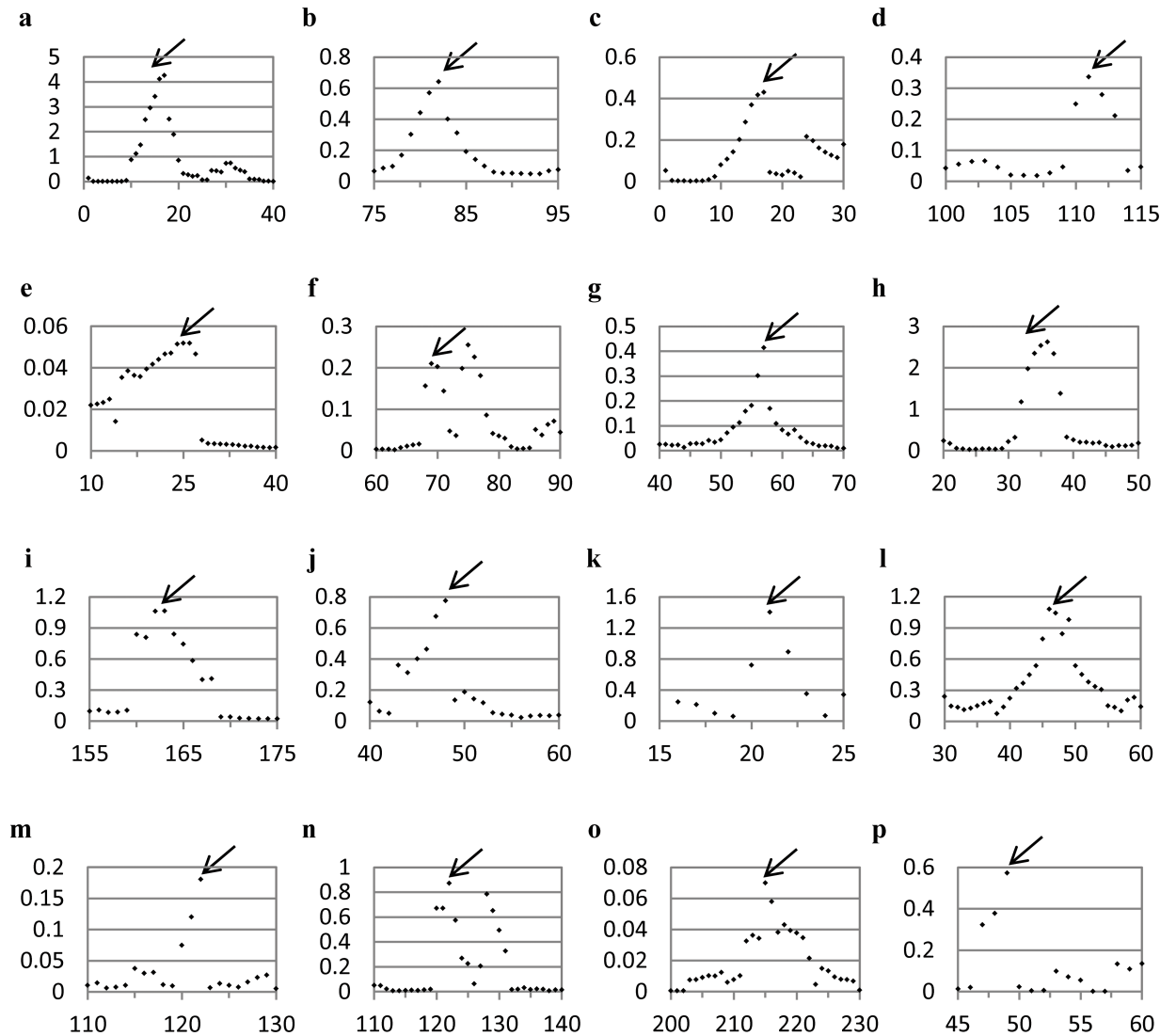


Fig 6. Additional running ruler symmetry analyses (see also S2 Fig); the neighborhood of the hinge region is shown. Black arrows—the hinge region. (a) Bovine seminal ribonuclease (PDB code: 1BSR), (b) β -crystallin (PDB code: 1BLB), (c) Human pancreatic ribonuclease chimera (PDB code: 1H8X), (d) RNase A N-terminal timer (PDB code: 1JS0), (e) Human glyoxalase I dimer (PDB code: 1BH5), (f) α -spectrin (PDB code: 2SPC), (g) Amyloid-like Cystatin C (PDB code: 1TIJ), (h) SH3 domain of Eps8 (PDB code: 1AOJ), (i) Circadian Clock Protein KaiA (PDB code: 1R8J), (j) Cyanovirin-N (PDB code: 1L5B), (k) Triggering receptor expressed on myeloid cells 1 (TREM-1) (PDB code: 1Q8M), (l) Cystatin A (PDB code: 1N9J), (m) Grb2-SH2 domain dimer (PDB code: 1FYR), (n) Odorant binding protein dimer (PDB code: 1OBP), (o) Cell division protein FtsZ (PDB code: 1W5F), (p) NrdH-redoxin (PDB code: 1R7H). See Table 1 for more information. For data sources see ref.'s [44–59].

<https://doi.org/10.1371/journal.pone.0180030.g006>

highest peak of interleukin-5 at 38–45 (Fig 5D, by a red arrow) belongs to a segment which is structurally located near the hinge region of the second monomeric sub-unit (Fig 7B). Since the hinges pair region itself is asymmetric, it exerts its distortive influence on neighboring surrounding areas by inter-segment interactions. These neighboring segments are loops, which are flexible areas, thus their distortion surpass that of the hinge-pair areas. The practical conclusion is that if one selects the symmetry analysis tool in order to identify possible hinge areas, then if several peaks appear in the spectrum, visual inspection, as is often practiced in reports on the domain-swapping mechanism, is helpful in eliminating non-relevant segments.

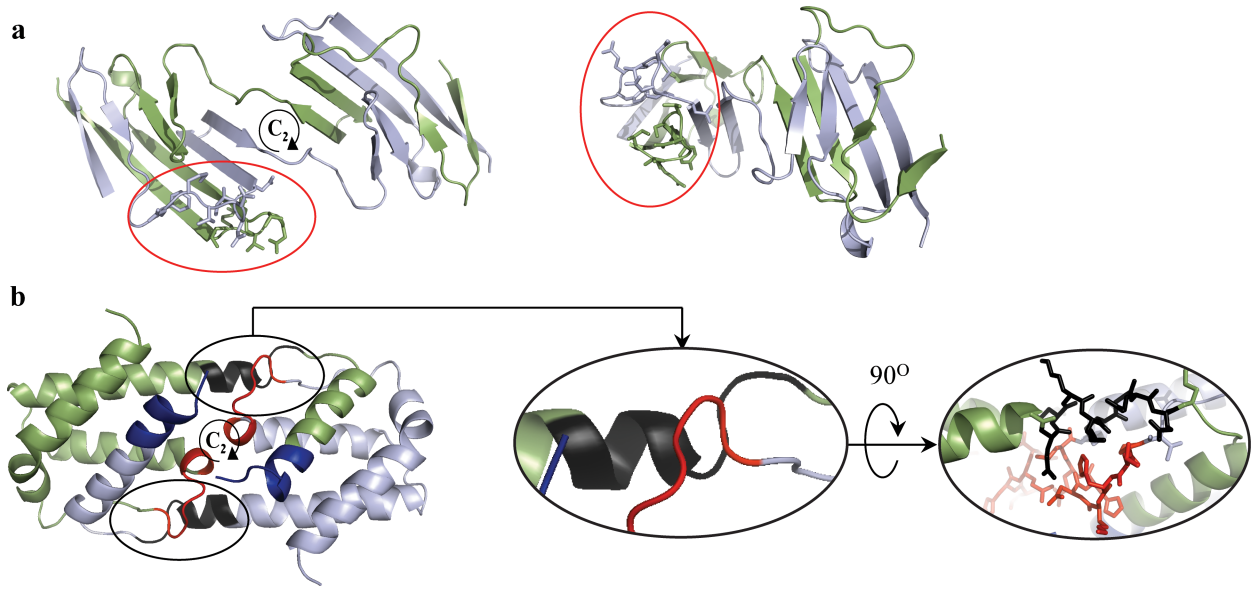


Fig 7. Focus on the origin of the symmetry distortion: each subunit is indicated by different color. (a) N-terminal domain of CD2 (PDB code: 1A64) from two different points of view. The amino-acids segments 21–28 and 81–88 are indicated by sticks representation. The interaction between those segments causes the symmetry distortion. These over-crowded regions are surrounded by red circles. (b) Interleukin-5 (PDB code: 1HUL). The regions which are indicated by colored arrows in Fig 5D are colored here respectively. The marked interaction area is seen from two different points of view. For data sources see ref.'s [41,42].

<https://doi.org/10.1371/journal.pone.0180030.g007>

Next, we demonstrate the usefulness of the symmetry analysis, when one wishes to analyze differences in propositions as to hinge identifications by various methods. For example, the reported proposition of Eisenberg[15] for the hinge area location in the TrkA-d4 dimer is the short segment of three amino-acids at positions 297–299. On the other and, Huang et al[19] used Eisenberg's method followed by manual inspection of the structure and proposed that the hinge area is wider and spans over positions 295–299. We have tested these two propositions by producing CSM spectra once with a running ruler of size 3, and once with size 5 (Fig 5E, black dots and red triangles, respectively). With size 3 (according to Eisenberg) the peak appears at 295, that is, the hinge region location is 295–297, a significant shift for such a small hinge region; however, when size 5 is applied (according to Huang) the spectrum indicates the location of the hinge region to be 295–299, in agreement with Huang et al. The fact that size 5 is apparently more relevant than size 3 is also in agreement with our previous analyzed example, drawing attention to the possibility that the distortive effect of the hinge is exerted beyond its minimal suggested size.

We also examined the possibility that the swapped-dimer hinge regions, which are the sites of maximal asymmetry, are also related to maximal flexibility. We therefore generated flexibility spectra for domain swapped structures in Figs 3 and 5 where hinge regions are the sites of maximal asymmetry. The flexibility of each segment in the spectrum was represented by the average atomic displacement factor (ADPs, crystallographic temperature factors) of the atoms in this segment, and the results are shown in S3 Fig. As can be seen, there is no correlation between the CSM spectrum and the flexibility spectrum of each protein. In each spectrum the hinge region is indicated by a local peak, and it is clearly seen that it is not the highest peak. This observation strengthens the interpretations provided by the CSM analysis tool, because it shows that the symmetry distortion of the hinge regions is not a thermal noise phenomenon.

Probability analysis

In this section we answer the following question: since the identification of the hinge region is based on the assumption that symmetry deviations tend to concentrate in that region, what is the probability that the observed hinge symmetry deviation is more than would be expected from random distribution of asymmetries throughout the protein? For that purpose we resort to the local symmetry analysis, which as explained in the Methods section, evaluates the CSM value of C_2 -symmetry related amino-acid pairs (one amino-acid in one monomer, and its counter near C_2 -symmetric amino-acid on the second monomer). In a sense, this analysis may also be considered as a "running ruler" analysis with a ruler size of one amino-acid. Here are the details of the statistical probability analysis:

We first run the local symmetry analysis on the whole protein, and get a list of all $S(C_2)$ values of all of the amino-acid pairs of the protein; that list is composed of N numbers, the number of amino-acids in one monomeric polypeptide chain in the oligomer. That list is arranged in a descending order of the $S(C_2)$ values, out of which the first d -most distorted pairs are taken, where d can be any number smaller than or equal to N ($d \leq N$). Next we check how many—(if any) of these d -most distorted pairs appear in the hinge of length h . We then evaluate the probability, $P(r)$, that $r = x$ distorted amino-acid pairs from the d -list will appear in a stretch of length h within a protein of length N . The probability that *at least* x amino-acids are in the hinge must include also the probability to find $r = x + 1$ amino acids from the d -list, $r = x + 2$ amino acids and so on, up to h amino-acids from the d -list. For our specific application, we find it therefore relevant to take the special case of $d = h$, for which $P(r)$ is:

$$P(r) = \frac{\binom{h}{r} \cdot \binom{N-h}{h-r}}{\binom{N}{h}}.$$

(See [S1 Appendix](#) and [S1 Fig](#) for the derivation of this equation). The probability that at least x amino-acids appear in the hinge of length h -length is then:

$$P = \sum_{r=x}^{r=h} P(r).$$

Applying this calculation we found ([Table 2](#)) that in the vast majority of the analyzed proteins, the number of the most distorted amino-acids which reside in the hinge exceeds by far the probability of that to happen, compared to random distribution of these distorted amino-acids in the whole protein. For instance, let us take again the RNase A N-terminal swapped dimer ([Fig 3](#), and PDB code 1A2W in [Table 2](#)), which has a hinge region size of $h = 8$ amino-acids. Five amino-acids in the protein are found in its hinge region, and thus $x = 5$. The calculated probability of that to happen coincidentally in a protein of 124 amino-acids (the size of each subunit) is 0.001%. It should be noted that the condition $d = h$ is quite stringent, because it may well be that the symmetry deviation of $d > h$ amino acids is considerable as well, and in that case the, the chances of having a symmetry-distorted amino-acid in the hinge, increases. Let us check for example $d = 2 \cdot h$ for the same RNase A N-terminal swapped dimer ($h = 8$). Increasing d to be 16 ($2 \cdot h$) changes the list of the most distorted amino acids to: 85, 22, 20, 101, 17, 98, 21, 19, 18, 100, 81, 16, 23, 99, 28, 31. This means that now all the amino acids in the hinge region (underlined) are in the list of the most distorted amino acids. The probability of that to happen coincidentally is $1 \cdot 10^{-6}\%$, namely three orders of magnitude less than the probability presented above.

Table 2. Probability analysis of symmetry distortion at the hinge range.

PDB ID	Polypeptide chain length	Hinge region	Most distorted amino-acids ^(a)	Quantity of most distorted amino-acids in hinge region	Probability of the observation
1BSR	124	17–24	<u>20</u> , <u>34</u> , <u>37</u> , <u>17</u> , <u>19</u> , <u>22</u> , <u>18</u> , <u>24</u>	6 / 8	5*10 ⁻³ %
1A2W	124	16–23	<u>85</u> , <u>22</u> , <u>20</u> , <u>101</u> , <u>17</u> , <u>98</u> , <u>21</u> , <u>19</u>	5 / 8	1*10 ⁻³ %
1BLB	174	81–89	<u>86</u> , <u>107</u> , <u>87</u> , <u>88</u> , <u>127</u> , <u>169</u> , <u>89</u> , <u>148</u> , <u>84</u>	5 / 9	1*10 ⁻³ %
1L5X	270	245–249	<u>247</u> , <u>245</u> , <u>38</u> , <u>248</u> , <u>43</u>	3 / 5	3*10 ⁻³ %
1QB3	109	89–95	<u>91</u> , <u>90</u> , <u>9</u> , <u>38</u> , <u>37</u> , <u>89</u> , <u>95</u>	4 / 7	0.02%
1CQZ	479	223–238	<u>94</u> , <u>233</u> , <u>135</u> , <u>231</u> , <u>91</u> , <u>90</u> , <u>188</u> , <u>232</u> , <u>223</u> , <u>93</u> , <u>427</u> , <u>204</u> , <u>430</u> , <u>420</u> , <u>63</u> , <u>228</u>	5 / 16	0.1%
2C5J	82	63–67	<u>66</u> , <u>63</u> , <u>6</u> , <u>64</u> , <u>69</u>	3 / 5	0.1%
1R8J	264	163–171	<u>170</u> , <u>259</u> , <u>167</u> , <u>271</u> , <u>238</u> , <u>169</u> , <u>232</u> , <u>231</u> , <u>278</u>	3 / 9	0.2%
1EN7	157	62–71	<u>67</u> , <u>71</u> , <u>89</u> , <u>68</u> , <u>93</u> , <u>87</u> , <u>70</u> , <u>28</u> , <u>96</u> , <u>74</u>	4 / 10	0.2%
2QYP	78	21–23	<u>23</u> , <u>21</u> , <u>19</u>	2 / 3	0.3%
2HKN	72	37–39	<u>38</u> , <u>36</u> , <u>37</u>	2 / 3	0.3%
1JS0	124	111–114	<u>113</u> , <u>64</u> , <u>112</u> , <u>69</u>	2 / 4	0.5%
2SPC	107	70–73	<u>1</u> , <u>3</u> , <u>73</u> , <u>71</u>	2 / 4	0.6%
1TIJ	112	57–60	<u>9</u> , <u>106</u> , <u>58</u> , <u>57</u>	2 / 4	0.6%
1CDC	96	43–49	<u>46</u> , <u>47</u> , <u>36</u> , <u>79</u> , <u>43</u> , <u>91</u> , <u>88</u>	3 / 7	0.8%
1AOJ	60	36–41	<u>39</u> , <u>21</u> , <u>38</u> , <u>37</u> , <u>20</u> , <u>19</u>	3 / 6	1%
2CN4	173	47–52	<u>39</u> , <u>49</u> , <u>74</u> , <u>51</u> , <u>63</u> , <u>88</u>	2 / 6	1%
1WWA	101	295–299	<u>296</u> , <u>347</u> , <u>349</u> , <u>295</u> , <u>350</u>	2 / 5	2%
2NZ7	93	91–95	<u>39</u> , <u>101</u> , <u>91</u> , <u>95</u> , <u>35</u>	2 / 5	2%
1WKQ	155	117–123	<u>60</u> , <u>61</u> , <u>76</u> , <u>119</u> , <u>123</u> , <u>96</u> , <u>148</u>	2 / 7	3%
1L5B	101	48–53	<u>49</u> , <u>37</u> , <u>52</u> , <u>1</u> , <u>40</u> , <u>32</u>	2 / 6	4%
1SCE	102	86–91	<u>60</u> , <u>88</u> , <u>87</u> , <u>35</u> , <u>101</u> , <u>59</u>	2 / 6	4%
1E7D	157	87–101	<u>68</u> , <u>74</u> , <u>93</u> , <u>89</u> , <u>96</u> , <u>101</u> , <u>67</u> , <u>71</u> , <u>59</u> , <u>22</u> , <u>28</u> , <u>98</u> , <u>38</u> , <u>43</u> , <u>27</u>	4 / 15	4%
1Q8M	121	21–23	<u>22</u> , <u>124</u> , <u>63</u>	1 / 3	7%
1A64	94	42–48	<u>44</u> , <u>27</u> , <u>87</u> , <u>26</u> , <u>86</u> , <u>45</u> , <u>20</u>	2 / 7	8%
1G6U	48	32–33	<u>32</u> , <u>42</u>	1 / 2	8%
1N9J	98	46–49	<u>2</u> , <u>48</u> , <u>1</u>	1 / 3	9%
1FYR	95	122–124	<u>122</u> , <u>135</u> , <u>145</u>	/ 3	9%
1OBP	155	121–124	<u>96</u> , <u>157</u> , <u>153</u> , <u>123</u>	1 / 4	10%
1LMK	239	122–126	<u>16</u> , <u>126</u> , <u>85</u> , <u>127</u> , <u>84</u>	1 / 5	10%
1W5F	316	215–220	<u>288</u> , <u>284</u> , <u>300</u> , <u>306</u> , <u>308</u> , <u>217</u>	1 / 6	11%
1R7H	74	49–51	<u>40</u> , <u>37</u> , <u>49</u>	1 / 3	12%
5CRO	61	54–56	<u>39</u> , <u>56</u> , <u>60</u>	1 / 3	14%
1XOG	104	34–37	<u>36</u> , <u>94</u> , <u>93</u> , <u>98</u>	1 / 4	15%
2OYA	102	428–431	<u>431</u> , <u>419</u> , <u>460</u> , <u>418</u>	1 / 4	15%
2OCT	95	46–49	<u>68</u> , <u>30</u> , <u>81</u> , <u>46</u>	1 / 4	16%
1BH5	177	24–36	<u>28</u> , <u>179</u> , <u>183</u> , <u>27</u> , <u>77</u> , <u>110</u> , <u>143</u> , <u>13</u> , <u>95</u> , <u>16</u> , <u>157</u> , <u>96</u>	2 / 13	24%
1J30	137	69–82	<u>22</u> , <u>80</u> , <u>15</u> , <u>41</u> , <u>47</u> , <u>138</u> , <u>134</u> , <u>135</u> , <u>108</u> , <u>139</u> , <u>120</u> , <u>51</u> , <u>3</u> , <u>69</u>	2 / 14	43.3%
1H8X	126	17–24	<u>125</u> , <u>17</u> , <u>31</u> , <u>66</u> , <u>91</u> , <u>100</u> , <u>37</u> , <u>67</u>	1 / 8	42%
1HUL	107	81–87	<u>110</u> , <u>112</u> , <u>50</u> , <u>109</u> , <u>42</u> , <u>40</u> , <u>111</u> , <u>39</u>	0 / 8	(see text)

^(a) Amino acids are listed by decreasing order; amino acids in the hinge region are underlined.

See Fig 3, Figs 5 and 6 and S2 Fig. For data sources see ref.'s [38–77].

<https://doi.org/10.1371/journal.pone.0180030.t002>

Returning to [Table 2](#), similar ($d = h$) calculations carried out on all of the proteins analyzed above ([Fig 3](#), [Figs 5 and 6](#) and [S2 Fig](#)), indicate that the probabilities of having the actual observed concentration of distortion in the hinge area, are all well below 15%. As exceptions are highlighting the rule, we comment on the last entry in the Table, Interleukin-5 (IL-5, PDB code: 1HUL): This protein does not have “most-distorted amino-acids” in the hinge region because another region in the protein is more distorted—see [Fig 5D](#)—and yet, as also seen in that figure, applying the running ruler analysis clearly identifies the whole hinge region as a peak in the CSM spectrum.

Conclusions

In conclusion, in relation to the question of ‘why do oligomers settle for imperfect symmetry if symmetrization is so advantageous’ we have explored here the parameter of the mechanism of the oligomerization. Taking the domain swapping mechanism we have shown that the mechanism of oligomerization is an important parameter in affecting the symmetry of the final oligomer (other key parameters are listed in the Introduction). The structure of protein oligomers is a reflection of their formation, and this is translated into the symmetry distortions. The new way of looking at swapped domain dimeric proteins offered by this study—through symmetry—allows comparative quantification of the effects of that mechanism. This method identifies the hinge regions in those proteins through the symmetry perspective, with no need of structural information on the monomeric form of the non-swapped protein (information that does not always existed). In many cases this symmetry analysis indicates the hinge segments as the major contributor to the symmetry distortions in the protein (it is always a contributor, even if not the major one). We found that in the vast majority of the analyzed proteins, the number of the most distorted amino-acids which reside in the hinge exceeds by far the probability of that to happen, compared to random distribution of these distorted amino-acids in the whole protein. And last but not least, we showed that the CSM spectrum and the running ruler method can be used generally for analyzing structural features of proteins, other than those associated with the hinge region.

Supporting information

S1 Appendix. Further explanation about the hinge symmetry probability analysis.
(PDF)

S1 Fig. Visual explanation of the probability calculation for the question ‘what is the probability that at least x amino-acids out of the d most distorted amino-acids appear in a given h -length-segment?’. The assumptions: (a) $N = 7$, namely, a dimeric protein composed of two subunits, each of 7-amino acids (a row of circles). (b) $h = 3$. The length of the hinge region is 3 amino acids and it placed as a sequence of at locations 2,3,4 (indicated by the bar); (c) $d = h = 3$. There is a list of the 3 most distorted amino acids (orange circles); (d) The experimental observation is that 2 out of the $d = 3$ most distorted amino-acids are located in the

hinge. There are $\binom{N}{d} = 35$ ways of placing the 3 most-distorted amino-acids in the set of 7 amino-acids; in each of these ways, the hinge region contains 0–3 amino acids out of the 3 most distorted amino-acids ($r = 0,1,2,3$).

(PDF)

S2 Fig. Additional running ruler symmetry analyses (see also [Fig 5](#)). The neighborhood of the hinge region is shown. Black arrows—the hinge region. (a) scaffold protein IscA (1XOG),

(b) sulerythrin (PDB code: 1J30), (c) Soluble epoxide hydrolase (PDB code: 1CQZ), (d) Cyclin-dependent kinase (PDB code: 1QB3), (e) Designed helical bundle (PDB code: 1G6U), (f) Endonuclease VII (PDB code: 1EN7), (g) Guanine deaminase (PDB code: 1WKQ), (h) T-SNARE (PDB code: 2C5J), (i) Hemophore HasA (PDB code: 2CN4), (j) Dynactin-1 (PDB code: 2HKN), (k) Caspase-recruitment domain CARD (PDB code: 2NZ7), (l) Cystatin B (PDB code: 2OCT), (m) Macrophage receptor MARCO (PDB code: 2OYA), (n) Saposin C Dimer (PDB code: 2QYP), (o) Survival protein E (PDB code: 1L5X), (p) Endonuclease VII (PDB code: 1E7D), (q) Suc1 (PDB code: 1SCE), (r) Cro repressor protein (PDB code: 5CRO). See [Table 1](#) for more information. For data sources see ref.'s [60–77].

S3 Fig. Comparison of the CSM running ruler symmetry analysis with the average atomic displacement factor (ADP) flexibility parameter. The black arrows indicate the hinge regions. PDB codes of analyzed proteins: (a) 1A2W, (b) 1CDC, (c) 1A64, (d) 1WWA.

Acknowledgments

Maayan Bonjack-Shterengartz is supported by the Ariane de Rothschild Women Doctoral Program. Useful discussions with Prof. Jonathan Breuer, Institute of Mathematics, The Hebrew University, are gratefully acknowledged. We thank Dr. Inbal Tuvi-Arad, The Open University, Israel, for useful advice.

Author Contributions

Conceptualization: Maayan Bonjack-Shterengartz, David Avnir.

Writing – original draft: Maayan Bonjack-Shterengartz.

Writing – review & editing: David Avnir.

References

1. Goodsell DS, Olson AJ. Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct.* 2000; 29: 105–153. <https://doi.org/10.1146/annurev.biophys.29.1.105> PMID: 10940245
2. André I, Strauss CEM, Kaplan DB, Bradley P, Baker D. Emergence of symmetry in homooligomeric biological assemblies. *Proc Natl Acad Sci U S A.* 2008; 105: 16148–16152. <https://doi.org/10.1073/pnas.0807576105> PMID: 18849473
3. Blundell TL, Srinivasan N. Symmetry, stability, and dynamics of multidomain and multicomponent protein systems. *Proc Natl Acad Sci.* 1996; 93: 14243–14248. PMID: 8962033
4. Kojić-Prodić B, Štefanić Z. Symmetry versus asymmetry in the molecules of life: Homomeric protein assemblies. *Symmetry (Basel).* 2010; 2: 884–906.
5. Berchanski A, Segal D, Eisenstein M. Modeling oligomers with Cn or Dn symmetry: Application to CAPRI target 10. *Proteins.* 2005; 60: 202–206. <https://doi.org/10.1002/prot.20558> PMID: 15981250
6. Taylor WR, May ACW, Brown NP, Aszodi A. Protein structure: geometry, topology and classification. *Reports Prog Phys.* 2001; 64: 517–590.
7. Taylor WR, Aszodi A. *Protein Geometry, Classification, Topology and Symmetry: A Computational Analysis of Structure (Series in Biophysics).* New York, USA: Taylor & Francis; 2004.
8. Venkatakrishnan AJ, Levy ED, Teichmann SA. Homomeric protein complexes: Evolution and assembly. *Biochem Soc Trans.* 2010; 38: 879–882. <https://doi.org/10.1042/BST0380879> PMID: 20658970
9. Marsh JA, Teichmann SA. Structure, dynamics, assembly, and evolution of protein complexes. *Annu Rev Biochem.* 2015; 84: 551–575. <https://doi.org/10.1146/annurev-biochem-060614-034142> PMID: 25494300
10. Bonjack-Shterengartz M, Avnir D. The near-symmetry of proteins. *Proteins.* 2015; 83: 722–734. <https://doi.org/10.1002/prot.24706> PMID: 25354765

11. Hurtley SM, Helenius A. Protein oligomerization in the endoplasmic reticulum. *Annu Rev Cell Biol.* 1989; 5: 277–307. <https://doi.org/10.1146/annurev.cb.05.110189.001425> PMID: 2688707
12. D'Alessio G. The evolutionary transition from monomeric to oligomeric proteins: tools, the environment, hypotheses. *Prog Biophys Mol Biol.* 1999; 72: 271–298. PMID: 10581971
13. Green SM, Gittis AG, Meeker AK, Lattman EE. One-step evolution of a dimer from a monomeric protein. *Nat Struct Biol.* 1995; 2: 746–751. PMID: 7552745
14. Bennett MJ, Schlunegger MP, Eisenberg D. 3D domain swapping: A mechanism for oligomer assembly. *Protein Sci.* 1995; 4: 2455–2468. <https://doi.org/10.1002/pro.5560041202> PMID: 8580836
15. Liu Y, Eisenberg D. 3D domain swapping: As domains continue to swap. *Protein Sci.* 2002; 11: 1285–1299. <https://doi.org/10.1110/ps.0201402> PMID: 12021428
16. Gronenborn AM. Protein acrobatics in pairs-dimerization via domain swapping. *Curr Opin Struct Biol.* 2009; 19: 39–49. <https://doi.org/10.1016/j.sbi.2008.12.002> PMID: 19162470
17. Liu S. A review on protein oligomerization process. *Int J Precis Eng Manuf.* 2015; 16: 2731–2760.
18. Liu C, Sawaya MR, Eisenberg D. β 2-microglobulin forms three-dimensional domain-swapped amyloid fibrils with disulfide linkages. *Nat Struct Mol Biol.* 2011; 18: 49–55. <https://doi.org/10.1038/nsmb.1948> PMID: 21131979
19. Huang Y, Cao H, Liu Z. Three-dimensional domain swapping in the protein structure space. *Proteins.* 2012; 80: 1610–1619. <https://doi.org/10.1002/prot.24055> PMID: 22411444
20. Chu CH, Lo WC, Wang HW, Hsu YC, Hwang JK, Lyu PC, et al. Detection and alignment of 3D domain swapping proteins using angle-distance image-based secondary structural matching techniques. *PLoS One.* 2010; 5: e13361. <https://doi.org/10.1371/journal.pone.0013361> PMID: 20976204
21. Shingate P, Sowdhamini R. Analysis of domain-swapped oligomers reveals local sequence preferences and structural imprints at the linker regions and swapped interfaces. *PLoS One.* 2012; 7: e39305. <https://doi.org/10.1371/journal.pone.0039305> PMID: 22848353
22. Rousseau F, Schymkowitz JWH, Itzhaki LS. The unfolding story of three-dimensional domain swapping. *Structure.* 2003; 11: 243–251. PMID: 12623012
23. Cámara-Artigas A. Crystallographic studies on protein misfolding: Domain swapping and amyloid formation in the SH3 domain. *Arch Biochem Biophys.* 2016; 602: 116–126. <https://doi.org/10.1016/j.abb.2016.02.024> PMID: 26924596
24. Lin YW, Nagao S, Zhang M, Shomura Y, Higuchi Y, Hirota S. Rational design of heterodimeric protein using domain swapping for myoglobin. *Angew Chem Int Ed Engl.* 2015; 54: 511–515. <https://doi.org/10.1002/anie.201409267> PMID: 25370865
25. Xu D, Tsai CJ, Nussinov R. Mechanism and evolution of protein dimerization. *Protein Sci.* 1998; 7: 533–544. <https://doi.org/10.1002/pro.5560070301> PMID: 9541384
26. Linhananta A, Zhou H, Zhou Y. The dual role of a loop with low loop contact distance in folding and domain swapping. *Protein Sci.* 2002; 11: 1695–1701. <https://doi.org/10.1110/ps.0205002> PMID: 12070322
27. Ding F, Prutzman KC, Campbell SL, Dokholyan NV. Topological determinants of protein domain swapping. *Structure.* 2006; 14: 5–14. <https://doi.org/10.1016/j.str.2005.09.008> PMID: 16407060
28. Levy Y, Cho SS, Shen T, Onuchic JN, Wolynes PG. Symmetry and frustration in protein energy landscapes: A near degeneracy resolves the Rop dimer-folding mystery. *Proc Natl Acad Sci U S A.* 2005; 102: 2373–2378. <https://doi.org/10.1073/pnas.0409572102> PMID: 15701699
29. Zabrodsky H, Peleg S, Avnir D. Continuous symmetry measures. *J Am Chem Soc.* 1992; 114: 7843–7851.
30. Dryzun C, Avnir D. Generalization of the continuous symmetry measure: The symmetry of vectors, matrices, operators and functions. *Phys Chem Chem Phys.* 2009; 11: 9653–9666. <https://doi.org/10.1039/b911179d> PMID: 19851543
31. Pinsky M, Dryzun C, Casanova D, Alemany P, Avnir D. Analytical methods for calculating Continuous Symmetry Measures and the Chirality Measure. *J Comput Chem.* 2008; 29: 2712–2721. <https://doi.org/10.1002/jcc.20990> PMID: 18484634
32. Salomon Y, Avnir D. Continuous symmetry measures: A note in proof of the folding/unfolding method. *J Math Chem.* 1999; 25: 295–308.
33. Pinsky M, Avnir D. Continuous Symmetry Measures. 5. The Classical Polyhedra. *Inorg Chem.* 1998; 37: 5575–5582. PMID: 11670704
34. Mezey PG. Fuzzy electron density fragments in macromolecular quantum chemistry, combinatorial quantum chemistry, functional group analysis, and shape–activity relations. *Acc Chem Res.* 2014; 47: 2821–2827. <https://doi.org/10.1021/ar5001154> PMID: 25019572

35. Lipiński PFJ, Dobrowolski JC. Local chirality measures in QSPR: IR and VCD spectroscopy. *RSC Adv.* 2014; 4: 47047–47055.
36. Shameer K, Shingate PN, Manjunath SCP, Karthika M, Pugalenth G, Sowdhamini R. 3DSwap: curated knowledgebase of proteins involved in 3D domain swapping. *Database* 2011; 2011: bar042. <https://doi.org/10.1093/database/bar042> PMID: 21959866
37. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28: 235–242. PMID: 10592235
38. Liu Y, Hart PJ, Schlunegger MP, Eisenberg D. The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-Å resolution. *Proc Natl Acad Sci U S A.* 1998; 95: 3437–3442. PMID: 9520384
39. Murray AJ, Lewis SJ, Barclay AN, Brady RL. One sequence, two folds: a metastable structure of CD2. *Proc Natl Acad Sci U S A.* 1995; 92: 7337–7341. PMID: 7638192
40. Perisic O, Webb PA, Holliger P, Winter G, Williams RL. Crystal structure of a diabody, a bivalent antibody fragment. *Structure.* 1994; 2: 1217–1226. PMID: 7704531
41. Murray AJ, Head JG, Barker JJ, Brady RL. Engineering an intertwined form of CD2 for stability and assembly. *Nat Struct Biol.* 1998; 5: 778–782. <https://doi.org/10.1038/1816> PMID: 9731771
42. Milburn M V., Hassell AM, Lambert MH, Jordan SR, Proudfoot AEI, Graber P, et al. A novel dimer configuration revealed by the crystal structure at 2.4 Å resolution of human interleukin-5. *Nature.* 1993; 363: 172–176. <https://doi.org/10.1038/363172a0> PMID: 8483502
43. Ultsch MH, Wiesmann C, Simmons LC, Henrich J, Yang M, Reilly D, et al. Crystal structures of the neurotrophin-binding domain of TrkA, TrkB and TrkC. *J Mol Biol.* 1999; 290: 149–159. <https://doi.org/10.1006/jmbi.1999.2816> PMID: 10388563
44. Mazzarella L, Capasso S, Demasi D, Di Lorenzo G, Mattia CA, Zagari A, et al. Bovine seminal ribonuclease: structure at 1.9 Å resolution. *Acta Crystallogr Sect D Biol Crystallogr.* 1993; 49: 389–402.
45. Nalini V, Bax B, Driessen H, Moss DS, Lindley PF, Slingsby C. Close packing of an oligomeric eye lens beta-crystallin induces loss of symmetry and ordering of sequence extensions. *J Mol Biol.* 1994; 236: 1250–1258. PMID: 8120900
46. Canals A, Pous J, Guasch A, Benito A, Ribó M, Vilanova M, et al. The structure of an engineered domain-swapped ribonuclease dimer and its implications for the evolution of proteins toward oligomerization. *Structure.* 2001; 9: 967–976. PMID: 11591351
47. Liu Y, Gotte G, Libonati M, Eisenberg D. Structures of the two 3D domain-swapped RNase A trimers. *Protein Sci.* 2002; 11: 371–380. <https://doi.org/10.1110/ps.36602> PMID: 11790847
48. Ridderström M, Cameron AD, Jones TA, Mannervik B. Involvement of an active-site Zn²⁺ ligand in the catalytic mechanism of human glyoxalase I. *J Biol Chem.* 1998; 273: 21623–21628. PMID: 9705294
49. Yan Y, Winograd E, Viel A, Cronin T, Harrison SC, Branton D. Crystal structure of the repetitive segments of spectrin. *Science.* 1993; 262: 2027–2030. PMID: 8266097
50. Janowski R, Kozak M, Abrahamson M, Grubb A, Jaskolski M. 3D domain-swapped human cystatin C with amyloidlike intermolecular β-sheets. *Proteins.* 2005; 61: 570–578. <https://doi.org/10.1002/prot.20633> PMID: 16170782
51. Kishan K V, Scita G, Wong WT, Di Fiore PP, Newcomer ME. The SH3 domain of Eps8 exists as a novel intertwined dimer. *Nat Struct Biol.* 1997; 4: 739–743. PMID: 9303002
52. Barrientos LG, Louis JM, Botos I, Mori T, Han Z, O’Keefe BR, et al. The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure.* 2002; 10: 673–686. PMID: 12015150
53. Ye S, Vakonakis I, Ioerger TR, LiWang AC, Sacchettini JC. Crystal structure of circadian clock protein KaiA from *Synechococcus elongatus*. *J Biol Chem.* 2004; 279: 20511–20518. <https://doi.org/10.1074/jbc.M400077200> PMID: 15007067
54. Radaev S, Kattah M, Rostro B, Colonna M, Sun PD. Crystal structure of the human myeloid cell activating receptor TREM-1. *Structure.* 2003; 11: 1527–1535. PMID: 14656437
55. Zerovnik E, Jerala R, Kroon-Zitko L, Turk V, Lohner K, Fierke C, et al. Characterization of the equilibrium intermediates in acid denaturation of human stefin B. *Eur J Biochem.* 1997; 245: 364–372. PMID: 9151965
56. Schiering N, Casale E, Caccia P, Giordano P, Battistini C. Dimer formation through domain swapping in the crystal structure of the Grb2-SH2-Ac-pYVNV complex. *Biochemistry.* 2000; 39: 13376–13382. PMID: 11063574
57. Tegoni M, Ramoni R, Bignetti E, Spinelli S, Cambillau C. Domain swapping creates a third putative combining site in bovine odorant binding protein dimer. *Nat Struct Biol.* 1996; 3: 863–867. PMID: 8836103
58. Oliva MA, Cordell SC, Löwe J. Structural insights into FtsZ protofilament formation. *Nat Struct Mol Biol.* 2004; 11: 1243–1250. <https://doi.org/10.1038/nsmb855> PMID: 15558053

59. Stehr M, Lindqvist Y. NrdH-redoxin of *Corynebacterium ammoniagenes* forms a domain-swapped dimer. *Proteins*. 2004; 55: 613–619. <https://doi.org/10.1002/prot.20126> PMID: 15103625
60. Mura C, Katz JE, Clarke SG, Eisenberg D. Structure and function of an archaeal homolog of survival protein E (SurEalpha): an acid phosphatase with purine nucleotide specificity. *J Mol Biol*. 2003; 326: 1559–1575. PMID: 12595266
61. Bourne Y, Watson MH, Arvai AS, Bernstein SL, Reed SI, Tainer JA. Crystal structure and mutational analysis of the *Saccharomyces cerevisiae* cell cycle regulatory protein Cks1: implications for domain swapping, anion binding and protein interactions. *Structure*. 2000; 8: 841–850. PMID: 10997903
62. Argiriadi MA, Morisseau C, Hammock BD, Christianson DW. Detoxification of environmental mutagens and carcinogens: structure, mechanism, and evolution of liver epoxide hydrolase. *Proc Natl Acad Sci U S A*. 1999; 96: 10637–10642. PMID: 10485878
63. Fridmann-Sirkis Y, Kent HM, Lewis MJ, Evans PR, Pelham HRB. Structural analysis of the interaction between the SNARE Tlg1 and Vps51. *Traffic*. 2006; 7: 182–190. <https://doi.org/10.1111/j.1600-0854.2005.00374.x> PMID: 16420526
64. Zhang G, Darst SA, Cotton R, Lilley D, Iwai S, Ohtsuka E, et al. Structure of the *Escherichia coli* RNA polymerase alpha subunit amino-terminal domain. *Science*. 1998; 281: 262–266. PMID: 9657722
65. Rossmann M, Schultz-Heienbrok R, Behlke J, Remmel N, Alings C, Sandhoff K, et al. Crystal structures of human saposins C and D: implications for lipid recognition and membrane interactions. *Structure*. 2008; 16: 809–817. <https://doi.org/10.1016/j.str.2008.02.016> PMID: 18462685
66. Honnappa S, Okhrimenko O, Jaussi R, Jawhari H, Jelesarov I, Winkler FK, et al. Key interaction modes of dynamic +TIP networks. *Mol Cell*. 2006; 23: 663–671. <https://doi.org/10.1016/j.molcel.2006.07.013> PMID: 16949363
67. Czjzek M, Létóffé S, Wandersman C, Delepierre M, Lecroisey A, Izadi-Pruneyre N. The crystal structure of the secreted dimeric form of the hemophore HasA reveals a domain swapping with an exchanged heme ligand. *J Mol Biol*. 2007; 365: 1176–1186. <https://doi.org/10.1016/j.jmb.2006.10.063> PMID: 17113104
68. Srimathi T, Robbins SL, Dubas RL, Hasegawa M, Inohara N, Park YC. Monomer/dimer transition of the caspase-recruitment domain of human Nod1. *Biochemistry*. 2008; 47: 1319–1325. <https://doi.org/10.1021/bi7016602> PMID: 18186648
69. Chagot B, Diochot S, Pimentel C, Lazdunski M, Darbon H. Solution structure of APETx1 from the sea anemone *Anthopleura elegantissima*: a new fold for an HERG toxin. *Proteins*. 2005; 59: 380–386. <https://doi.org/10.1002/prot.20425> PMID: 15726634
70. Bourne Y, Arvai AS, Bernstein SL, Watson MH, Reed SI, Endicott JE, et al. Crystal structure of the cell cycle-regulatory protein suc1 reveals a beta-hinge conformational switch. *Proc Natl Acad Sci U S A*. 1995; 92: 10232–10236. PMID: 7479758
71. Raaijmakers H, Törö I, Birkenbihl R, Kemper B, Suck D. Conformational flexibility in T4 endonuclease VII revealed by crystallography: implications for substrate binding and cleavage. *J Mol Biol*. 2001; 308: 311–323. <https://doi.org/10.1006/jmbi.2001.4592> PMID: 11327769
72. Ogihara NL, Ghirlanda G, Bryson JW, Gingery M, DeGrado WF, Eisenberg D. Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc Natl Acad Sci U S A*. 2001; 98: 1404–1409. <https://doi.org/10.1073/pnas.98.4.1404> PMID: 11171963
73. Ohlendorf DH, Tronrud DE, Matthews BW. Refined structure of Cro repressor protein from bacteriophage λ suggests both flexibility and plasticity. *J Mol Biol*. 1998; 280: 129–136. <https://doi.org/10.1006/jmbi.1998.1849> PMID: 9653036
74. Morimoto K, Yamashita E, Kondou Y, Lee SJ, Arisaka F, Tsukihara T, et al. The asymmetric IscA homodimer with an exposed [2Fe-2S] cluster suggests the structural basis of the Fe-S cluster biosynthetic scaffold. *J Mol Biol*. 2006; 360: 117–132. <https://doi.org/10.1016/j.jmb.2006.04.067> PMID: 16730357
75. Ojala JRM, Pikkarainen T, Tuuttila A, Sandalova T, Tryggvason K. Crystal structure of the cysteine-rich domain of scavenger receptor MARCO reveals the presence of a basic and an acidic cluster that both contribute to ligand recognition. *J Biol Chem*. 2007; 282: 16654–16666. <https://doi.org/10.1074/jbc.M701750200> PMID: 17405873
76. Jenko Kokalj S, Gunčar G, Štern I, Morgan G, Rabzelj S, Kenig M, et al. Essential Role of Proline Isomerization in Stefin B Tetramer Formation. *J Mol Biol*. 2007; 366: 1569–1579. <https://doi.org/10.1016/j.jmb.2006.12.025> PMID: 17217964
77. Fushinobu S, Shoun H, Wakagi T. Crystal structure of sulerythrin, a rubrerythrin-like protein from a strictly aerobic archaeon, *Sulfolobus tokodaii* strain 7, shows unexpected domain Swapping. *Biochemistry*. 2003; 42: 11707–11715. <https://doi.org/10.1021/bi034220b> PMID: 14529281